

STATS 4CI3/6CI3 Winter 2021

ASSIGNMENT 3

Submit to Crowdmark using the link that was emailed to you.

Due before 11 PM on Wednesday, February 24th.

Your assignment must conform to the Assignment Standards listed below.

Assignments submitted up to 24 hours late will incur a 30% penalty.
Assignments submitted more than 24 hours late will receive a zero grade.

Answer all questions, stating your answer and showing your code. Not all questions carry equal marks.

Set a seed where appropriate to make your work reproducible

1. (15 MARKS) Compute the integral of

$$g(x) = (\cos(30x) + \sin(10x))^2$$

between 0 and 1,

- (a) deterministically, by using the `integrate()` function in R.
- (b) approximately, using Monte Carlo Integration with sample size 10,000.

Hint: $\int_0^1 g(x)dx = \int_0^1 g(x)f(x)dx$, where $f(x) = 1$ is the uniform density.

2. (15 MARKS)

- (a) Predict the number of girls in 600 births, where the probability of female birth is 51.1%
- (b) Repeat the simulation 1000 times and produce a histogram to show the distribution of results.

3. **(15 MARKS)**

Simulate, using a simulation size of 1000, from the following linear model,

$$y = 0.5 + 2x + \epsilon$$

where $\epsilon \sim N(0, 2^2)$. Assume $x \sim N(0, 1)$.

- (a) display the results in a scatterplot of x against y
- (b) report the median value of y

4. **(25 MARKS)**

Use 100,000 replications when calculating the probabilities of both (A) and (B) below.

Which is more likely (include the calculated probabilities in your answer):

- (A) getting at least one 6 when rolling a single fair six-sided die 4 times or
- (B) getting at least one pair of sixes when two fair, six-sided dice are thrown 24 times.

Hint: you need to find a way to write the desired condition using logical operations and the `sum` function.

5. **(30 MARKS)**

In an experiment, the null hypothesis is defined as $H_0 : \mu = 100$.

Obtain a sample of 25 observations from a Normal distribution with mean 100 and standard deviation 15. Using this sample, run a t-test using the `t.test` function in order to test the null hypothesis, (using a p-value significance level of .05), and

- (a) report the resulting p-value
- (b) report the test-statistic
- (c) do you reject or fail to reject the null hypothesis?

Repeat the experiment 10,000 times (i.e. generate 10,000 p-values) and

- (d) generate a histogram to illustrate the distribution of the p-values obtained
- (e) state what proportion of your results were statistically significant

Assignment Standards

- L^AT_EX is strongly recommended but not strictly required. The use of Markdown in R studio is also recommended.
 - Submit your assignment as one **.pdf document**. **All R code should be included inline.**
 - Do not include a title page. The title, your **name and student number** should be printed at the top of the first page.
 - The writing and referencing should be appropriate to the university level.
 - Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.
-